DOCUMENT RESUME

ED 426 098 TM 029 328

AUTHOR	Frederick, Brigitte N.
TITLE	Fixed-, Random-, and Mixed-Effects ANOVA Models: A User-Friendly Guide for Increasing the Generalizability of ANOVA Results.
PUB DATE	1999-01-00
NOTE	23p.; Paper presented at the Annual Meeting of the Southwest Educational Research Association (San Antonio, TX, January 21-23, 1999).
PUB TYPE	Guides - Non-Classroom (055) -- Speeches/Meeting Papers (150)
EDRS PRICE	MF01/PC01 Plus Postage.
DESCRIPTORS	*Analysis of Variance; *Mathematical Models; Research Design
IDENTIFIERS	*Fixed Effects; Generality; Random Variables; *Rules of Thumb (Statistics)

ABSTRACT

Most researchers using analysis of variance (ANOVA) use a fixed-effects model. However, a random- or mixed-effects model may be a more appropriate fit for many research designs. One benefit of the random- and mixed-effects models is that they yield more generalizable results. This paper focuses on the similarities and differences between the various ANOVA models and the factors that should be considered when determining which model to use. An explanation of the "Rules of Thumb" for deriving the correct formulas for computing "F" statistics is also given. (Contains 2 tables and 15 references.) (SLD)

Fixed-, Random-, and Mixed-Effects ANOVA Models: A User-Friendly Guide For Increasing the Generalizability of ANOVA Results

Brigitte N. Frederick

Texas A & M University

Paper presented at the annual meeting of the Southwest Educational Research Association, San Antonio, January, 1999.

Abstract

Most researchers using ANOVA procedures employ a fixed-effects model. However, a random- or mixed-effects model may be a more appropriate fit for many research designs. One benefit of random- and mixed-effects models is that they yield more generalizable results. This paper focuses on the similarities and differences between the various ANOVA models and the factors that should be considered when determining which model to utilize. A description of the "Rules of Thumb" for deriving the correct formulas for computing $\underline{F}$ statistics will also be explained.

Fixed-, Random-, and Mixed-Effects ANOVA Models: A User-Friendly Guide For Increasing the Generalizability of ANOVA Results

Most researchers using analysis of variance (ANOVA) procedures choose a *fixed-effects* model, although they may not realize that they are making this choice or realize its consequences. Although fixed-effects models are the most common of the ANOVA models, they are not necessarily the most appropriate and/or useful procedures for all experimental designs. There are two additional types of ANOVA models that are less commonly used: random-effects (also called Model II) models and mixed-effects (also called Model III) models.

A *random-effects* model is used when the researcher wants to generalize findings to levels/conditions within ways/factors beyond the levels that are represented in the study (Jackson & Brashers, 1994). Hays (1981) commented, "the random-effects model is designed especially for experiments in which inferences are to be drawn about an entire set of distinct treatments or factor levels, including some not actually observed" (p. 376). A *mixed-effects* model is a combination of fixed- and random-effects models, comprised of one or more random ways and one or more fixed ways.

The purpose of the present paper is to outline the differences between the fixed-, mixed- and random-effects ANOVA designs. Tips on deciding whether a way should be classified as fixed or random will also be provided. Computational differences, as well as the "rules of thumb"

for deriving the correct formulas to compute F statistics in the random and mixed models will also be explained.

Three Models

In a fixed-effect model, the null hypothesis is that there are no differences between the means of the levels of the way that are utilized in the study (Ostle, 1988). **In contrast, the hypothesis tested in a random-effects model is that there are no differences in the means of all of the levels of a way that are possible in the population of levels that the sample of levels was chosen from, not just in the sample of levels that is utilized in the study** (Ostle, 1988). Thus, random-effects models attempt to generalize beyond both the sampled people and the sampled levels.

There is some incongruence between what researchers consider a random way. Some argue that any way that does not include all the possible levels should be treated as random (Clark, 1973; Richter & Seay, 1987). Others disagree. Wike and Church (1976) outline three basic methods for selecting the levels of a way to be used in a study. First, the researcher could use all of the possible levels of a way (number of levels, *p*, EQUALS the number of levels in the population of levels, *P*). Clearly, this is rarely the case in actual research. Second, the researcher could randomly choose a subset of all possible levels of a way to use in the study (number of levels, *p*, is less than the number of levels in

the population, *P*, with random selection of levels). Third, the researcher could choose to nonrandomly select a subset of all of the possible levels of a way to use in the study (number of levels in the sample, *p*, is less than the number of levels in the population, *P*, with nonrandom selection of levels). It seems clear that the p=P case must be treated as a fixed-effects model. In the p<P case, random selection implies a random-effects model (Wike & Church, 1976). Mathematical and statistical formulas typically treat the third case (p<P, nonrandom selection case) as fixed (Wike & Church, 1976). However, some researchers argue that the third case should also be treated as a random-effects model (Clark, 1973; Richter & Seay, 1987).

This debate is clearly illustrated by Clark (1973) in his article, "Language-as-Fixed Effect Fallacy: A Critique of Language Statistics in Psychological Research." Using the example of language research, Clark argued that words, as well as participants, should be treated as random effects because the words that are selected for language studies do not extinguish all possibilities within the entire population of words. Therefore, if the researcher wishes to generalize results beyond the scope of the study the ways must be classified as random. Clark (1973) commented,

> When should the investigator treat language as a random effect? The answer is, whenever the language stimuli do not deplete the population from which

> they were drawn. The answer is *not*, whenever the language stimuli used were chosen at *random* from this population. (p. 348)

Richter and Seay (1987) agreed that classifying ways as random provides statistical support for further generalization of results. They replicated a study on recognition memory and found that when words were considered fixed there were three effects statistically significant at the .05 level and four statistically significant at the .01 level. However, when words were reclassified as random only one of the original seven effects remained statistically significant. Table 1 presents these results. Although all studies will not have findings that differ this dramatically across models, the example illustrates the differences that are possible when ways for the same data are classified as random instead of fixed. In response to Clark's article, Wike and Church (1976) argued that while classifying ways as random can lead to a decrease in Type I errors, *overclassifying* ways as random can lead to an increase in Type II errors. They also contended that generalization can not be assumed just because ways are classified as random. They noted, "generality is not obtained simply by selecting *p* levels randomly" (Wike & Church, 1976, p. 253).

One method to determine if a way would best be classified as fixed or random is to see if the levels are "interchangeable" (Jackson & Brashers, 1994; Ostle, 1988).

That is, could alternate levels of the way be substituted in the study without making a difference in the results? With a random way, a set of levels could be utilized on one run of the experiment and an entirely different set could be used on a subsequent replication (Glass & Hopkins, 1984). **In a random-effects model the specific levels of a way chosen to be utilized in a study are of no particular interest to the researcher. Therefore, the levels can be substituted without changing the fundamental research question.**

For example, a researcher could be studying the effectiveness of three teaching methods: peer tutoring, computer assisted instruction, and lecturing. Instead of taking samples of students from every grade kindergarten through twelve, the researcher could randomly select a subset of grades to be present in the study. The researcher could then generalize the results of the study to all of the grades that are possible, even those not randomly selected and present in the study. It would make no difference to the researcher if the particular grades in the study were 1st, 2nd, 3rd, 5th, and 8th or 2nd, 4th, 7th, 10th, and 12th; the grades represented in the study are of **no particular interest** to the researcher. However, if the researcher were to use different kinds of instruction (for example, reading from the text or parent tutoring) at each selected grade level, the study would have an altered meaning. Therefore, this example would be a mixed study, where the grade of the

students is a random way and the mode of instruction is a fixed way.

A second method for determining if a way should be classified as random or fixed is based on what conclusions the researcher would like to draw from the results of the study (Jackson & Brashers, 1994). If the researcher wants to make conclusions based only on the levels of the population that are used in the study it is appropriate to classify a way as fixed. However, if the researcher would like to draw conclusions about populations beyond the levels that are represented in the study, the way must be classified as random. Using the teaching method example to demonstrate this point, think of the conclusions that can be drawn if the grade way is treated as fixed. It could be concluded, for example, that peer tutoring is better than computer assisted instruction and lecturing for individuals in grades 2, 4, 7, 10, and 12, if these were the grade levels actually studied. In contrast, a conclusion that could be drawn when the grade way is classified as random would be that peer tutoring is better than computer assisted instruction and lecturing for students in *all* grades.

Depending on the desires of the researcher and the research question, the same ways can be classified differently in various studies. The determination of whether a way should be classified as random or fixed ultimately depends on the context of the research (Longford, 1993). However, the determination of whether a way will be

considered random or fixed must be made prior to the data collection and analysis. Hicks (1973) noted,

> It is not reasonable to decide after the data have been collected whether the levels are to be considered fixed or random. This decision must be made prior to the running of the experiment, and if random levels are to be used, they must be chosen from all possible levels by a random process. (p. 173)

What are the repercussions of misclassifying a way? Of course, there are no *absolutely right* answers to how a way should be classified. Coleman (1979) stated, "there are usually several correct ways to analyze an experiment, and...the better choice is more a matter of wisdom than mathematical correctness" (p. 243). However, there are always situations where it is *most appropriate* to classify a way in a particular way. In these situations if a way is misclassified, different undesirable consequences can occur.

If a fixed way is misclassified as random, it will be subject to a overly conservative test of statistical significance and therefore the likelihood of making a Type II error (not rejecting a false null hypothesis) will increase (Wike & Church, 1976). Inversely, if a random way is misclassified as fixed, there is a greater chance of making a Type I error (falsely rejecting a true null hypothesis) (Clark, 1973). In addition, if a random way is classified as

fixed, the results of the study can not be generalized beyond the levels that are utilized in the study (Clark, 1973).

There are several assumptions that go along with the random-effects model. First, it is assumed that the levels of the random ways in the experiment represent only a random sample of all of the possible levels (Mason, Gunst & Hess, 1989). Second, the random variables representing main effects and interactions, as well as the error terms, are assumed to be statistically independent of one another (Mason et al., 1989). Third, the main effect, interaction, and error terms are assumed to be normally distributed with a mean of zero (Mason et al., 1989). Fourth, the ANOVA assumption of homogeneity of variance is presumed, with the added component of *sphericity*, which means that the variances of the differences in means must be homogeneous across pairs of levels in the study (Jackson & Brashers, 1994). As in fixed way ANOVAs, the more that these assumptions are met, the more accurate the findings of the study are (Jackson & Brashers, 1994). In addition to these assumptions, the random levels of a random or mixed design must have equal Ns, not as a matter of ease, but because no methods have been fully developed for the unbalanced random- or mixed-effect design (Glass & Hopkins, 1984).

Different Denominators

When doing computations for random- and mixed-effects models, the *same* general procedures are used as for a fixed-

effects ANOVA, with the *exception* of the F ratio test statistic. The sources of variance, degrees of freedom, sum of squares, and mean squares are computed *identically* to the fixed-effects model (Hinkle, Wiersma, & Jurs, 1998). The difference in computations is in finding the appropriate terms for the F ratio. Whereas the denominator of the F ratio is always the mean squares error for the fixed-effects models, the *denominator* varies in the mixed- and random-effects models (Hinkle et al., 1998).

Why is it necessary to use a different error term in the denominator of the F ratio for designs incorporating random ways? According to Kennedy and Bush (1985), error variance in a fixed model is theoretically drawn from one source: variance between participants (also called SOS within, residual, unexplained). However, in a random model there are two sources of error variance: variance between participants (within) and variance from randomly sampling from all possible levels. Therefore, to find the true effect for a way it is necessary to divide by a more complex term than MS error to remove via the division the effect of the multiple sources of variance.

"Rules of Thumb" for Determining the Correct F Denominator

The Expected Mean Squares [E(MS)] is the statistic that determines what the terms in the F ratio test statistic should be in a random- or mixed-effects model (Hinkle et al., 1998). The E(MS) can be defined as the hypothetical mean

value of a variance computed over infinitely many observations of an experiment (Jackson & Brashers, 1994). Although the calculations for the E(MS) are incredibly difficult and complicated, there are some "rules of thumb" that can help to estimate the E(MS) (Hicks, 1973; Ott, 1984).

To illustrate the procedure, we will go back to the previous example with three methods of teaching five different grades of children. The design utilized in the study will be a 3 x 5 mixed model with a fixed A way (method of teaching) and a random B way (grade of the student). For heuristic purposes, we will establish an N of 30 (two per cell). The "rules of thumb" involve five steps.

1. List sources of variance in the study as row labels in a two-way table.

Main effect A_i
Main effect B_j
Interaction AB_{ij}
Error $E_{k(ij)}$

2. Write the subscript listed in each row label as a column heading. Be sure that every subscript appears only once (denoted "A" in the chart below). In the column headings also write R or F to signify if the effect is random or fixed ("B"). A term is considered random if ANY of the ways that contribute to it are random, so all interactions involving one or more random ways are random. Last, in each column

heading write the number of observations associated with the way. For the effects, this will be the number of levels; for the error term this will be the number of replications or people in each cell ("C").

(C) ⇒	**5**	**3**	**2**
(B) ⇒	**R**	**F**	**R**
(A) ⇒	**i**	**j**	**k**
Main effect A_i			
Main effect B_j			
Interaction AB_{ij}			
Error $E_{k(ij)}$			

3. Transfer the number in the top row of the column heading to each of the rows in that column that the subscript does <u>not</u> appear in the row label.

	5	3	2
	R	F	R
	i	j	k
Main effect A_i		**3**	**2**
Main effect B_j	**5**		**2**
Interaction AB_{ij}			**2**
Error $E_{k(ij)}$			

4. For any subscripts that are in parentheses in the row labels of the chart, place a one in the columns that are headed by the letter(s) in the brackets.

	5	3	2
	R	F	R
	i	j	k
Main effect A_i		3	2
Main effect B_j	5		2
Interaction AB_{ij}			2
Error $E_{k(ij)}$	**1**	**1**	

5. Fill the remaining cells in the table with zeros if the column is for a fixed way or ones if the column is for a random way.

	5	3	2
	R	F	R
	i	j	k
Main effect A_i	**0**	3	2
Main effect B_j	5	**1**	2
Interaction AB_{ij}	**0**	**1**	2
Error $E_{k(ij)}$	1	1	**1**

The completed "rules of thumb" table can then be used to compute the E(MS) values. To do this, cover the column(s) that the letter in the top row if the column heading matches the subscript in the row label. For example, for main effect A you would cover the first column because the subscript is *i*; for the interaction effect you would cover the first two columns because it has the subscripts *i* and *j*. Next, determine which rows are going to be used for computing the E(MS) of a particular way by looking at the subscripts for each row. If the subscript of the way that you are computing appears in the subscript for another row, you will use the row in your E(MS) computations. For example, when computing the E(MS) for main effect A you would use the A way main effect, the AB interaction, and the error variance rows because the subscript *i* appears in each row. You would not use the B way main effect because the subscript for the row is *j*. Third, for all of the rows that are being used in the computation, multiply the numbers in each of the rows that are not covered up and then multiply this number by the variance term that is associated with the row. For the random and mixed ways, the variance term will be σ^2. For the fixed ways, the variance term will be ϕ. The E(MS) for each way

will be the sum of these constant x variance term products. The E(MS) equation for the A way main effect is:

(3 x 2)(ϕ) from the A way main effect + (1 x 2)(σ^2) from the AB interaction effect + (1 x 1)(σ^2) from error

This would yield an E(MS) of $6\phi_A + 2\sigma_{AB}^2 + \sigma_E^2$.

	5 R i	3 F j	2 R k	**E(MS)**
Main effect A_i	**0**	3	2	$6\phi_A$* + $2\sigma_{AB}^2 + \sigma_E^2$
Main effect B_j	5	**1**	2	$10\sigma_B^2 + \sigma_E^2$
Interaction AB_{ij}	**0**	**1**	2	$2\sigma_{AB}^2 + \sigma_E^2$
Error $E_{k(ij)}$	1	1	**1**	σ_E^2

* The symbol ϕ signifies a fixed variance.

What does the E(MS) tell you? As you can see from the "rules of thumb" computations, the E(MS) takes into consideration the many sources of variation that are associated with the different effects in the model. For example, the A main effect E(MS) is:

$$6\phi_A + 2\sigma_{AB}^2 + \sigma_E^2$$

The first term in the equation signifies the variance that comes from the A way. The second term in the equation is the variance from the interaction effect. The third term in the equation is the error variance. In order to isolate the variance that is *only* associated with the A way main effect, we need to remove the effect of the interaction variance and the error variance from the A main effect E(MS) term (Hinkle et al., 1998). Looking back at the E(MS) in the "rules of thumb" chart, the term that has both the AB interaction variance and the error variance is the interaction E(MS), which is:

$$2\sigma_{AB}^2 + \sigma_E^2$$

If we divide the A main effect E(MS) by the interaction effect E(MS) to get the F ratio, we are left with only the variance that is associated with the A way. Therefore, we now have used the correct term for the denominator of the F ratio to test the A way main effect.

The same procedure can be used for the remaining effects in the model, giving us F ratios calculated by:

Main effect B: $10\sigma_B^2 + \sigma_E^2 / \sigma_E^2$

Interaction effect: $2\sigma_{AB}^2 + \sigma_E^2 / \sigma_E^2$

Table 2 illustrates the F ratios that are computed when different models are utilized.

Discussion

One crucial difference between fixed-effects models and random- and mixed-effects models is the use of planned and post hoc tests (Thompson, 1994). In a fixed model, once there is evidence of a statistically significant difference between means in a way, it is common practice to run a post hoc analysis to determine exactly where the differences exist. Alternatively, planned contrast analyses if the researcher has an informed guess about where the differences may be before the analysis is run. However, with a random effects model, it would make no sense to run an analysis of any type to determine where the differences are within the various levels of a way because the ways in the study are

arbitrarily and randomly selected. The researcher is not primarily concerned with where differences may be within a certain way because the levels are of no particular interest to the researcher. Kennedy and Bush (1985) commented,

> Since the levels of the random variable have not been deliberately chosen for subject-matter reasons, but rather have been selected because they are somewhat representative of a larger population of levels, it makes little sense to attempt specific comparisons among levels. (p. 294)

In conclusion, there are several benefits to using a mixed-effects or random-effects model, *if doing so is appropriate for the design of the study*. As in any study, decisions must be made based on the context of the experiment. Researchers are encouraged to contemplate how the levels utilized in their study are selected and what generalizations are desired from the study when choosing an ANOVA model. As was outlined by this paper, a random- or mixed-effects model may be more appropriate for many experiments that are generally analyzed as fixed-effects models, fixed-effects models may be over used because some statistical packages only provide these tests, or default to these tests. Although choosing to use a random- or mixed-effects model may require more work on the part of the

researcher, the increased generalizability of results makes it worth the effort.

References

Clark, H.H. (1973). The language-as-fixed-effect fallacy: A critique of language statistics in psychological research. Journal of Verbal Learning and Verbal Behavior, 12, 335-359.

Coleman, E.B. (1979). Generalization effects vs random effect: Is σ_{TL}^2 a source of type 1 or type 2 error? Journal of Verbal Learning and Verbal Behavior, 18, 243-256.

Glass, G.V. & Hopkins, K.D. (1984). Statistical methods in education and psychology (2nd ed.). Englewood Cliffs, NJ: Prentice-Hall.

Hays, W.L.(1981). Statistics (3rd ed.). New York: Holt, Rinehart, and Winston.

Hicks, C.R. (1973). Fundamental concepts in the design of experiments. New York: Holt, Rinehart, and Winston.

Hinkle, D.E., Wiersma, W., & Jurs, S.G. (1998). Applied statistics for the behavioral sciences. Boston: Houghton Mifflin.

Jackson, S., & Brashers, D.E. (1994). Random ways in ANOVA. Thousand Oaks, CA: SAGE.

Kennedy, J.J., & Bush, A.J. (1985). An introduction to the design and analysis of experiments in behavioral research. Lanham, MD: University Press of America.

Longford, N.L. (1993). Random coefficient models. New York: Oxford University Press.

Mason, R.L., Gunst, R.F., & Hess, J.L. (1989). Statistical design and analysis of experiments. New York: Wiley.

Ostle, B. (1988). Statistics in research: Basic concepts and techniques for research workers (4th ed.). Ames, IA: Iowa State Press.

Ott, L. (1984). Introduction to statistical methods and data analysis. Boston: Prindle, Weber, & Schmidt.

Richter, M.L., & Seay, M.B. (1987). ANOVA designs with subjects and stimuli as random effects: Applications to prototype effects on recognition memory. Journal of Personality and Social Psychology, 53(3), 470-480.

Thompson, B. (1994). Planned versus unplanned and orthogonal versus nonorthogonal contrasts: The neo-classic perspective. In B. Thompson (Ed.), Advances in social science methodology (Vol. 3, pp. 3-27). Greenwich, CT: JAI Press.

Wike, E.L., & Church, J.D.(1976). Comments on Clark's "The language-as-fixed-effect fallacy." Journal of Verbal Learning and Verbal Behavior, 15, 249-255.

Table 1

Differences in F Ratios For Fixed- and Random- Effects Models

Source	Fixed model F ratio	Random model F ratio
Explicit-Implicit(A)	1.650	1.582
Extrovert-Introvert(B)	4.398*	.388
Stereotype-Control(C)	28.519**	3.061
Word Categories(D)	3.106*	.171
A x B	3.187	2.406
A x C	1.561	1.366
A x D	.259	.539
B x C	34.373**	2.618
B x D	.812	.078
C x D	27.905**	1.899
A x B x C	.296	.483
A x B x D	1.419	1.372
A x C x D	2.021	1.313
B x C x D	4.690*	.307
A x B x C x D	.099	.368
Words (W) within BCD	27.511**	27.511**
A x W	1.200	1.200

*p<.05
**p<.01

Note. Adapted from Richter and Seay (1987, p. 475).

Table 2

Calculated F ratios for a two-way ANOVA

	Fixed Effects Model	Random Effects Model (Both ways random)	Mixed Effects Model (A way random)
A main effect	effect/ error	effect/ interaction	effect/ error
B main effect	effect/ error	effect/ interaction	effect/ interaction
Interaction effect	effect/ error	effect/ error	effect/ error

Note. Adapted from Hinkle et al. (1998).

TM029328

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement (OERI)
Educational Resources Information Center (ERIC)

REPRODUCTION RELEASE

(Specific Document)

ERIC

I. DOCUMENT IDENTIFICATION:

Title: FIXED-, RANDOM-, AND MIXED-EFFECTS ANOVA MODELS: A USER-FRIENDLY GUIDE FOR INCREASING THE GENERALIZABILITY OF ANOVA RESULTS	
Author(s): BRIGITTE N. FREDERICK	
Corporate Source:	Publication Date: 1/99

II. REPRODUCTION RELEASE:

In order to disseminate as widely as possible timely and significant materials of interest to the educational community, documents announced in the monthly abstract journal of the ERIC system, *Resources in Education* (RIE), are usually made available to users in microfiche, reproduced paper copy, and electronic/optical media, and sold through the ERIC Document Reproduction Service (EDRS) or other ERIC vendors. Credit is given to the source of each document, and, if reproduction release is granted, one of the following notices is affixed to the document.

If permission is granted to reproduce the identified document, please CHECK ONE of the following options and sign the release below.

[X] Sample sticker to be affixed to document	Sample sticker to be affixed to document [ ]
Check here Permitting microfiche (4"x 6" film), paper copy, electronic, and optical media reproduction — "PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY BRIGITTE N. FREDERICK TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."	"PERMISSION TO REPRODUCE THIS MATERIAL IN OTHER THAN PAPER COPY HAS BEEN GRANTED BY Sample TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)." — **or here** Permitting reproduction in other than paper copy.
Level 1	**Level 2**

Sign Here, Please

Documents will be processed as indicated provided reproduction quality permits. If permission to reproduce is granted, but neither box is checked, documents will be processed at Level 1.

"I hereby grant to the Educational Resources Information Center (ERIC) nonexclusive permission to reproduce this document as indicated above. Reproduction from the ERIC microfiche or electronic/optical media by persons other than ERIC employees and its system contractors requires permission from the copyright holder. Exception is made for non-profit reproduction by libraries and other service agencies to satisfy information needs of educators in response to discrete inquiries."

Signature: X Brigitte Fr	Position: RES ASSOCIATE
Printed Name:	Organization: TEXAS A&M UNIVERSITY
Address: TAMU DEPT EDUC PSYC COLLEGE STATION, TX 77843-4225	Telephone Number: (409) 845-1831
	Date: 11/9/98